ordinalToMetric(clusterSim)

Reinforcing measurement scale for ordinal data (ordinal to metric scale)

The GDM2 distance measure can be applied to reinforcing measurement scale for ordinal data.

1. We start with data matrix $[x_{ij}]$, where x_{ij} denotes *i*-th observation on *j*-th variable.

Table 1. Data matrix (27 residential housing properties described by six variables) – see `data_patternGDM2`

No.	x1	x2	x3	x4	x5	x6
1	5	3	1	3	1	3
2	3	3	3	3	2	2
3	5	4	3	4	1	2
4	2	3	1	3	2	3
5	5	4	2	4	1	2
6	4	3	2	3	1	3
7	3	4	3	3	2	2
8	4	4	3	4	1	1
9	5	3	2	4	1	2
10	4	2	1	3	1	3
11	5	4	3	4	1	4
12	4	3	1	4	1	2
13	4	4	3	3	1	1
14	4	4	3	3	2	3
15	5	4	2	3	2	4
16	3	3	2	3	1	1
17	4	2	1	3	2	3
18	4	1	2	4	1	2
19	3	3	2	3	2	4
20	3	2	1	3	1	3
21	4	3	2	3	1	1
22	5	3	2	4	1	2
23	5	4	3	4	1	2
24	4	2	2	3	1	2
25	3	2	1	2	2	3
26	3	3	1	1	2	3
27	2	3	1	1	2	3

Residential housing properties were described by the following ordinal variables:

x1 Location of environmental land, which is linked to a dwelling (1 – poor, 2 – inadequate, 3 – satisfactory, 4 – good, 5 – very good).

x2 Standard utility of a dwelling (1 – bad, 2 – low, 3 – average, 4 – high).

x3 Living conditions occurring on the land, which is linked to a dwelling (1 – bad, 2 – average, 3 – good).

x4 Location of land, which is related to dwelling in the area of the city (1 – central, 2 – downtown, 3 – intermediate, 4 – peripheral).

x5 Type of condominium (1 – low, 2 – large).

x6 Area of land, which is related to dwelling (1 – below the contour of the building, 2 – outline of the building, 3 – the outline of the building with the environment acceptable, such as parking, playground, 4 – the outline of the building with the environment too much).

2. Due to the fact that the method of reinforcing scale of ordinal data using GDM2 distance (see function `dist.GDM` in `clusterSim` package) will concern each variable separately, so the formula for the GDM2 distance (see distGDM_details.pdf) will be as follows:

$$d_{iw} = \frac{1}{2} - \frac{a_{iwj}b_{wij} + \sum_{\substack{l=1 \\ l\neq i,k}}^{n} a_{ilj}b_{wlj}}{2\left[\sum_{l=1}^{n} a_{ilj}^2 \cdot \sum_{l=1}^{n} b_{wlj}^2\right]^{\frac{1}{2}}} \text{ for } j = 1, \ldots, m, \tag{1}$$

where: $d_{iw} \in [0; 1]$ – GDM2 distance between *i*-th object and the pattern object *w*,
$i, k, l = 1, \ldots, n$ – the number of object,
$j = 1, \ldots, m$ – the number of variable,
x_{ij} (x_{kj}, x_{lj}) – *i*-th (*k*-th, *l*-th) observation on *j*-th variable,

$$a_{ipj}(b_{wrj}) = \begin{cases} 1 & \text{if} \quad x_{ij} > x_{pj}(x_{wj} > x_{rj}) \\ 0 & \text{if} \quad x_{ij} = x_{pj}(x_{wj} = x_{rj}) \\ -1 & \text{if} \quad x_{ij} < x_{pj}(x_{wj} < x_{rj}) \end{cases} \text{ for } p = w, l;\ r = i, l.$$

To transform the ordinal variable into the metric variable we use formula:

$$s_{iw} = 1 - d_{iw} \text{ for } j = 1, \ldots, m. \tag{2}$$

As a result of equation (2) will be transformation of ordinal variables into metric variables:

$$\text{Ordinal variable} \begin{bmatrix} x_{1j} \\ \vdots \\ x_{ij} \\ \vdots \\ x_{nj} \end{bmatrix} \Rightarrow \quad \begin{array}{c} \text{Counting GDM2 distances} \\ \text{between each } i\text{-th } (i = 1, \ldots, n) \\ \text{object and the pattern object } w \end{array} \quad \begin{bmatrix} s_{1j} \\ \vdots \\ s_{ij} \\ \vdots \\ s_{nj} \end{bmatrix} \Rightarrow \text{metric variable}$$

3a. The main goal of the linear ordering methods is to identify the order of the objects with respect to predetermined criterion. Usually the synthetic measure, which aggregates the partial information contained in the variables, is used.

Three types of performance variables are distinguished: stimulants – where higher value means better performance, destimulants – where low values indicate better performance, nominants – where the best value is implied. Object performance is positively assessed if the measure has implied value.

In this situation, in the formula (1) x_{wj} $(j = 1, \ldots, m)$ will indicate the most preferred category of all of the variable categories (maximal category for stimulants, minimal category for destimulants, nominal category for nominants).

The use of formula (2) will reinforce scale of ordinal variables to metric variables. Additionally all variables, which have nominant and destimulant character of type, are transformed into stimulant variables via formula (2).

Types of performance variables in Table 1:
x1, x2, x3 – stimulants,
x4, x5 – destimulants,
x6 – nominant (the nominal category: 3).

The coordinates of pattern object consist of the best variables' values: `patternCoordinates=c(5,4,3,1,1,3)`.

```
# Example 1
library(clusterSim)
data(data_patternGDM2)
res1<-ordinalToMetric(data_patternGDM2,scaleType="o",
      patternCoordinates=c(5,4,3,1,1,3))
print(res1)
```

3b. For some of multivariate statistical methods (cluster analysis, multidimensional scaling, factor analysis) in the formula (1) x_{wj} $(j = 1, \ldots, m)$ will indicate maximal category for each ordinal variable. The use of formula (2) will reinforce scale of ordinal variables to metric variables.

The coordinates of pattern object consist of the maximum variables' values: `patternCoordinates=c(5,4,3,4,2,4)`.

```
# Example 2
library(clusterSim)
data(data_patternGDM2)
res2<-ordinalToMetric(data_patternGDM2,scaleType="o",
      patternCoordinates=c(5,4,3,4,2,4))
print(res2)
```

References

Jajuga, K., Walesiak, M., Bak, A. (2003), *On the general distance measure*, In: M. Schwaiger, O. Opitz (Eds.), *Exploratory data analysis in empirical research*, Springer-Verlag, Berlin, Heidelberg, 104-109. DOI: https://doi.org/10.1007/978-3-642-55721-7_12

Walesiak, M. (1993), *Statystyczna analiza wielowymiarowa w badaniach marketingowych* [*Multivariate Statistical Analysis in Marketing Research*], Wroclaw University of Economics, Research Papers no. 654.

Walesiak, M. (1999), *Distance Measure for Ordinal Data*, Argumenta Oeconomica, No. 2 (8), 167-173.

Walesiak, M. (2011), *Uogólniona miara odległości GDM w statystycznej analizie wielowymiarowej z wykorzystaniem programu R* [*The Generalized Distance Measure GDM in multivariate statistical analysis with R]*, Wydawnictwo Uniwersytetu Ekonomicznego, Wroclaw.

Walesiak, M. (2014), *Wzmacnianie skali pomiaru dla danych porządkowych w statystycznej analizie wielowymiarowej* [*Reinforcing measurement scale for ordinal data in multivariate statistical analysis*], Prace Naukowe Uniwersytetu Ekonomicznego we Wroclawiu, Taksonomia 22.